

# Distributed Detection over Noisy Networks: Large Deviations Analysis

Dušan Jakovetić, José M. F. Moura, João Xavier

## Abstract

We study the large deviations performance of consensus+innovations distributed detection over *noisy* networks, where sensors at a time step  $k$  cooperate with immediate neighbors (consensus) and assimilate their new observations (innovation.) We show that, even under noisy communication, *all sensors* can achieve exponential decay  $e^{-kC_{\text{dis}}}$  of the detection error probability, even when certain (or most) sensors cannot detect the event of interest in isolation. We achieve this by designing a single time scale stochastic approximation type distributed detector with the optimal weight sequence  $\{\alpha_k\}$ , by which sensors weigh their neighbors' messages. The optimal design of  $\{\alpha_k\}$  balances the opposing effects of communication noise and information flow from neighbors: larger, slowly decaying  $\alpha_k$  improves information flow but injects more communication noise. Further, we quantify the best achievable  $C_{\text{dis}}$  as a function of the sensing signal and noise, communication noise, and network connectivity. Finally, we find a threshold on the communication noise power below which a sensor that can detect the event in isolation still improves its detection by cooperation through noisy links.

**Keywords:** Noisy communication, distributed detection, Chernoff information, large deviations.

The work of the first, and the third authors is partially supported by: the Carnegie Mellon—Portugal Program under a grant from the Fundação para a Ciência e Tecnologia (FCT) from Portugal; by FCT grants CMU-PT/SIA/0026/2009; and by ISR/IST plurianual funding (POSC program, FEDER). The work of J. M. F. Moura is partially supported by NSF under grants CCF-1011903 and CCF-1018509, and by AFOSR grant FA95501010291. D. Jakovetić holds a fellowship from FCT.

D. Jakovetić is with the Institute for Systems and Robotics (ISR), Instituto Superior Técnico (IST), Technical University of Lisbon, Lisbon, Portugal, and with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA [djakovet@andrew.cmu.edu](mailto:djakovet@andrew.cmu.edu)

J. M. F. Moura is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA [moura@ece.cmu.edu](mailto:moura@ece.cmu.edu)

J. Xavier is with the Institute for Systems and Robotics (ISR), Instituto Superior Técnico (IST), Technical University of Lisbon, Lisbon, Portugal [jxavier@isr.ist.utl.pt](mailto:jxavier@isr.ist.utl.pt)

## I. INTRODUCTION

Consider a generic network of sensors that sense the environment to detect an event of interest. In centralized detection, the measurements of *all* sensors at *all* times  $k$  are available at the detector (fusion center.) Under appropriate conditions, the probability of error  $P^e(k)$  of the centralized minimum probability of error detector decays in  $k$  at an exponential rate,  $P^e(k) \sim e^{-Ck}$ , where  $C$  is the (centralized) Chernoff information. We research in this paper the equivalent question of exponential rate of decay of the probability of error  $P^e$  for *distributed* detection at each local sensor. We consider this question when the (local) communications among sensors is through *noisy* links.

To be specific, we study consensus+innovations distributed algorithms like for example the LMS and RLS adaptive algorithms in [1], [2], [3], the detectors in [4], [5], or the estimators in [6]. In consensus+innovations distributed algorithms, at time  $k$ , each sensor updates its state 1) by a weighted average of the states of its neighbors (consensus); and 2) by incorporating its local measurement (innovations):

$$\text{state}_{k+1} = \text{state}_k + \gamma_k^1 \text{consensus}_k + \gamma_k^2 \text{innovations}_k. \quad (1)$$

Consensus+innovations detectors like in (1) are distributed, stochastic approximation type algorithms, but particular algorithms make different choices of the *time-decaying weight sequences*  $\gamma_k^i$ ,  $i = 1, 2$ , in (1) by which sensors weigh the consensus (their neighbors' messages) and the innovations (their own measurements) terms at each time  $k$ : [1], [2], [3] set  $\gamma_k^i = \mu$ ,  $i = 1, 2$ ; in [4] they vanish at the same rate; while [5], [6] consider single but also mixed scale algorithms where these weight sequences vanish at different rates. We will show that key to achieving exponential decay of the distributed detector  $P^e$  at *all* sensors is the suitable design of the weights  $\gamma_k^i$ ,  $i = 1, 2$ , in (1).

This paper addresses three natural questions:

- 1) *Centralized versus distributed (through noisy links) detection*: under which, if any, conditions can consensus+innovations distributed detection achieve *exponentially fast* decay of the (detection) error probability *at each sensor*,  $P^e \sim e^{-C_{\text{dis}}k}$ —the best possible decay rate  $C_{\text{dis}}$ , not necessarily equal to the centralized Chernoff information  $C$ ; in other words, when the sensors cooperate through noisy links, can distributed detection achieve at every sensor an exponential rate of decay like centralized detection. We answer affirmatively this question, under mild structural conditions, by careful design of the weight sequences in (1).
- 2) *Can cooperation through noisy links help*: How close can  $C_{\text{dis}}$  approach the corresponding (centralized) Chernoff information  $C$ ? We solve this by designing the *best* weight sequences  $\gamma_k^i$ ,  $i = 1, 2$ ,

that maximize  $C_{\text{dis}}$ . We explicitly quantify the optimal  $C_{\text{dis}}^*$  as a function of a *sensing* signal-to-noise ratio **SSNR** and a *communication* signal-to-noise ratio **CSNR** that we will define, and the network algebraic connectivity<sup>1</sup>. Our analysis reveals opposing effects: small and fast decaying weight  $\gamma_k^1$  injects less communication noise, but also reduces the information flow from neighbors; the optimal weights strike the best balance between these two effects.

- 3) **SSNR versus CSNR**—*how much communications noise can distributed detection sustain*: What is the highest communications noise level, i.e., lowest **CSNR**, for which cooperation helps? Let sensor  $i$  be the *best* sensor among all locally detectable sensors and assume that, without cooperation, its  $P^e \sim e^{-k C_i}$ . Can cooperation over noisy links make the worst sensor under communication better than  $i$ —the best one without cooperation? We explicitly find a *threshold* on the ratio **CSNR/SSNR** above which communication pays off in the latter sense; the threshold is a function of the network algebraic connectivity.

**Brief comment on the literature.** *Consensus+innovations* distributed algorithms as in (1) or in [4], [1], [2], [3], [5], [6], [8], that interleave *consensus* and *innovations* at the same time instant contrast with decentralized parallel fusion architectures, e.g., [9], [10], [11], [12], [13], where all sensors communicate with a fusion sensor or with consensus-based detection schemes (no fusion sensor,) for example, [14], [15], [16], [17], where sensors in the network, initially, 1) collect a single snapshot of measurements, and, *subsequently*, 2) run the consensus algorithm to fuse their decision rules.

We consider noisy communications among sensors. Communications imperfections in consensus-based detection in sensor networks are usually modeled via intermittent link failures and additive noise [18]. For consensus+innovations distributed algorithms, the LMS and RLS adaptive algorithms in [1], [2], [3] and the distributed change detection algorithm in [8] do not consider link failures nor additive noise. References [4], [19], [20] consider link failures but no additive communication noise. Reference [21] considers deterministically time varying networks. Reference [6] is concerned with estimation and considers a very general model that includes sensor failures, link failures, and various degrees of either quantized or noisy communications. To the best of our knowledge and within the consensus+innovations detectors, only reference [5] and now this paper consider additive noise in the communications among sensors, but no link failures. We highlight the main differences between our work and [5].

Reference [5] proposes a consensus+innovations distributed detector that it refers to as  $\mathcal{MD}$ . Algorithm

<sup>1</sup>The algebraic connectivity is the second smallest eigenvalue of the graph Laplacian matrix [7] that measures the speed of averaging across the network; larger algebraic connectivity means faster averaging.

$\mathcal{MD}$  assumes very general data distributions: temporally independent, spatially correlated sensing noise and temporally independent, spatially correlated additive communication noise, both with generic distributions with finite second moments. Under global detectability and connectedness assumptions, [5] shows that  $\mathcal{MD}$ 's error probability  $P^e$  decays to zero at all nodes  $i$ , but [5] only shows exponential decay rate of the error probability for a modified,  $\mathcal{SD}$  scheme, when the noises are Gaussian, and *all sensors are locally detectable, with equal Chernoff informations  $C_i > 0^2$*  ([5], Corollary 12.) In fact, as we show in this paper, Appendix B, the probability of error  $P^e$  for  $\mathcal{MD}$  is not exponential; it is instead *sub exponential*<sup>3</sup>, i.e., the rate is strictly slower than exponential, when the  $C_i$ 's are not equal, with possibly some  $C_i$ 's equal to zero. The subexponential rate of the  $\mathcal{MD}$  and  $\mathcal{SD}$  algorithms (with unequal  $C_i$ 's) is due to the decay rates assumed by these algorithms for the stochastic approximation weight sequences. In contrast, in the consensus+innovations algorithm that we propose, we craft carefully these weight sequences; this enables us to show for the Gaussian problem and under global detectability and connectedness that our distributed detector achieves exponential decay rate for  $P^e$  at every sensor, regardless of the equal or unequal  $C_i$ 's, where some can possibly be zero. Further, we optimize the weight sequences so that sensors achieve the maximum payoff from their (noisy) cooperation with other sensors. We derive our results on the  $P^e$  under Gaussian assumptions on the sensing and communication noises, but our results extend, to a certain degree, to the *non-Gaussian* (time-independent and space-independent) zero mean sensing noise with finite second moment and to the *non-Gaussian* (time-independent and space-correlated) zero mean communication noise with finite second moment.

The Gaussian assumptions allow us to completely characterize the rate of decay of the error probability solely on the basis of *the first two moments* (mean and variance) of the node's decision variable or state, say  $x_i(k)$ . With non-Gaussian noises, the first two moments no longer suffice to determine the rate of decay of the error probability  $P^e$ , but they still represent a good measure of detection performance. In this case, we can still show that the (local) detector signal-to noise ratio  $\mathbf{DSNR}_i(k)$  at each sensor  $i$  that we define by the ratio of the square of the mean over the variance of the sensor  $i$  state  $x_i(k)$  grows at the same rate  $\sim k$ , as for the optimal centralized detector.

We relate now this paper to our prior work on distributed detection, [19], [20]. While [19], [20] study the effect of link failures on detection performance, this paper addresses *additive communication noise* in the links among sensors. Our analysis here reveals that communication noise has an effect

<sup>2</sup>Sensor  $i$  can detect the event individually (is locally detectable) if and only if  $C_i > 0$ ; see ahead Definition 1 and Fact 2.

<sup>3</sup>We show in Appendix B that  $\max_{i=1,\dots,N} P_i^e(k)$ —the worst error error probability at time  $k$  among all  $N$  sensors, is at least  $e^{-ck^\tau}$ , where  $\tau \in (0.5, 1)$  and  $c > 0$ .

that is qualitatively different from that of link failures; with link failures, the more communication that is actually achieved among sensors the better the error performance, since when communication does happen sensors receive their neighbors decision variables unencumbered by noise. On the other hand, communication noise leads to a clear tradeoff between communication noise and information flow (degree at which consensus helps), with a cooperation payoff–threshold on the **CSNR**. To show these results, the analysis we develop here is very different from the analysis we advanced in [19], [20]. In [19], [20], we consider independent identically distributed (i.i.d.) averaging matrices  $W(k)$  (and hence the distribution of the  $W(k)$  is time invariant,) and no communication noise. In contrast, this paper considers time-decaying stochastic approximation weights (and hence, time varying weight matrices  $W(k)$ ) and additive communication noise; these additional challenges do not allow for our tools in [19], [20] and demand new analysis. A final comment to distinguish our methods here with those in [5]. Reference [5] uses standard stochastic approximation techniques [22] that yield the exact asymptotic covariance of the decision variable vector (when the local Chernoff informations are equal) the asymptotic covariance is given by a difficult to interpret matrix integration formula. In contrast, we do not pursue the exact asymptotic covariance of the decision variable, but get, instead, tight, simple, easy to interpret lower and upper bounds by exploiting the natural separability between the communication noise and the information flow (averaging) effects.

**Paper organization.** The next paragraph introduces notation. Section II describes the problem model and presents our distributed detector. Section III states our modeling assumptions and gives preliminary analysis. Section IV presents our main results on the asymptotic performance of our distributed detector. Section V proves our main results. Section VI presents extensions to the non-Gaussian case. Finally, section VII concludes the paper. Appendices A–B provide remaining proofs.

**Notation.** We denote by:  $A_{ij}$  or  $[A]_{ij}$  (as appropriate) the  $(i, j)$ -th entry of a matrix  $A$ ;  $a_i$  or  $[a]_i$  the  $i$ -th entry of a vector  $a$ ;  $A^\top$  the transpose of  $A$ ;  $I$ ,  $1$ , and  $e_i$ , respectively, the identity matrix, the column vector with unit entries, and the  $i$ -th column of  $I$ ,  $J := (1/N)11^\top$  the  $N \times N$  ideal averaging matrix;  $\|\cdot\|_l$  the vector (respectively, matrix)  $l$ -norm;  $\|\cdot\| = \|\cdot\|_2$  the Euclidean (respectively, spectral) norm;  $\lambda_i(\cdot)$  and  $\text{tr}(\cdot)$  the  $i$ -th smallest eigenvalue, and the trace of a matrix;  $\text{Diag}(a)$  the diagonal matrix with the diagonal equal to the vector  $a$ ;  $|\mathcal{A}|$  the cardinality of  $\mathcal{A}$ ;  $\mathbb{E}[\cdot]$ ,  $\text{Var}(\cdot)$ ,  $\text{Cov}(\cdot)$ , and  $\mathbb{P}(\cdot)$  the expected value, the variance, the covariance, and probability operators, respectively;  $\mathcal{I}_{\mathcal{A}}$  the indicator function of the event  $\mathcal{A}$ ;  $\mathcal{N}(\mu, \Sigma)$  the normal distribution with mean  $\mu$  and covariance  $\Sigma$ ;  $\mathcal{Q}(\cdot)$  the Q-function, i.e.,

the function that calculates the right tail probability of the standard normal distribution;

$$\mathcal{Q}(t) = \frac{1}{\sqrt{2\pi}} \int_t^{+\infty} e^{-\frac{u^2}{2}} du, \quad t \in \mathbb{R}. \quad (2)$$

We also make use of the standard  $\Omega$  and  $O$  notations:  $f(k) = \Omega(g(k))$  stands for existence of a  $K > 0$  such that  $f(k) \geq cg(k)$ , for some  $c > 0$ , for all  $k \geq K$ ; and  $f(k) = O(g(k))$  means existence of  $K > 0$  such that  $f(k) \leq cg(k)$ , for some  $c > 0$ , for all  $k \geq K$ .

## II. BINARY HYPOTHESES TESTING: CENTRALIZED AND DISTRIBUTED

This section presents the network model and our consensus+innovations distributed detector whose performance we analyze in section IV. The current section also considers the centralized and isolated sensor detectors for benchmarking our consensus+innovations detector and defines certain relevant signal-to-noise ratios.

### A. Network

We consider a network of  $N$  sensors. The topology of the network defines who can communicate with whom and is described by a simple (no self or multiple links,) undirected graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , where  $\mathcal{V}$  is the set of sensors with  $|\mathcal{V}| = N$ , and  $\mathcal{E}$  is the set of links or communication channels among sensors: the link between sensors  $i$  and  $j$  is represented in the graph by  $(i, j) \in \mathcal{E}$ . The neighborhood set  $O_i$  and the degree  $d_i$  of sensor  $i$  are  $O_i = \{j : (i, j) \in \mathcal{E}\}$  and  $d_i = |O_i|$ , respectively. The  $N \times N$  adjacency matrix is  $A = [A_{ij}]$ , with  $A_{ij} = 1$  if  $(i, j) \in \mathcal{E}$  and  $A_{ij} = 0$  else (with  $A_{ii} = 0$ , for all  $i$ .) The graph Laplacian is  $\mathcal{L} = D - A$ , where  $D = \text{Diag}(d_1, \dots, d_N)$ . Consider the eigenvalue decomposition

$$\mathcal{L} = Q \Lambda(\mathcal{L}) Q^\top, \quad (3)$$

where  $\Lambda(\mathcal{L}) = \text{Diag}(\lambda_1(\mathcal{L}), \dots, \lambda_N(\mathcal{L}))$ , with the eigenvalues in increasing order, and the columns  $q_i$  of  $Q$  are the orthonormal eigenvectors of  $\mathcal{L}$ . It is well known that  $\lambda_1(\mathcal{L}) = 0$  and  $q_1 = \frac{1}{\sqrt{N}} \mathbf{1}$ . Further,  $\mathcal{G}$  is connected if and only if the algebraic connectivity  $\lambda_2(\mathcal{L}) > 0$ , [7].

### B. Isolated sensor detector and centralized detector

We consider the known signal in Gaussian noise binary hypotheses test between  $H_1$  and  $H_0$ . At time  $k$ , sensor  $i$  measures the (scalar)  $y_i(k)$ :

$$\text{under } H_l : y_i(k) = [m_l]_i + \zeta_i(k), \quad l = 0, 1,$$

with prior probabilities  $0 < P(H_1) = 1 - P(H_0) < 1$ . Here  $[m_l]_i$  is a constant known signal and the *sensing* noise  $\{\zeta_i(k)\}$  is a zero mean (z.m.) independent identically distributed (i.i.d.) Gaussian sequence. Introduce the vector notation

$$y(k) = (y_1(k), \dots, y_N(k))^T, m_l = ([m_l]_1, \dots, [m_l]_N)^T, \zeta(k) = (\zeta_1(k), \dots, \zeta_N(k))^T.$$

The covariance of the *sensing* noise is  $S_\zeta = \text{Cov}(\zeta(k))$ .

**Isolated sensor detector.** A sensor working in isolation processes only its own observation. The test statistic is the local likelihood ratio ( $\ell$ LLR)  $\mathcal{D}_i(k)$ ,  $k = 1, \dots$ , given by the sum of the instantaneous  $\ell$ LLR  $L_i(j)$ ,  $j = 1, \dots, k$ , where

$$\begin{aligned} L_i(j) &= [m_1 - m_0]_i \frac{y_i(j) - \frac{[m_1]_i + [m_0]_i}{2}}{[S_\zeta]_{ii}} \\ \mathcal{D}_i(k) &= \frac{1}{k} \sum_{j=1}^k L_i(j). \end{aligned} \quad (4)$$

The isolated sensor  $i$  detector thresholds  $\mathcal{D}_i(k)$  against a threshold  $\tau_i(k)$ .

**Centralized detector.** The centralized log-likelihood ratio (cLLR) for the single vector measurement  $y(k)$  (all sensors measurements are available at the fusion center) is:

$$L(k) = (m_1 - m_0)^T S_\zeta^{-1} \left( y(k) - \frac{m_1 + m_0}{2} \right).$$

The cLLR at time  $k$  is:

$$\mathcal{D}(k) = \frac{1}{k} \sum_{j=1}^k L(j) \quad (5)$$

The optimal centralized detector thresholds the cLLR against  $\tau(k)$ . For future reference we introduce:

$$\eta(k) = (\eta_1(k), \eta_2(k), \dots, \eta_N(k))^T \quad (6)$$

$$\eta_i(k) = [S_\zeta^{-1}(m_1 - m_0)]_i \left( y_i(k) - \frac{[m_1]_i + [m_0]_i}{2} \right). \quad (7)$$

Conditioned on  $H_l$ ,  $l = 0, 1$ , the sequence  $\eta(k)$  is i.i.d. Gaussian with mean  $m_\eta^{(l)}$  and covariance  $S_\eta$ :

$$m_\eta^{(l)} = (-1)^{(l+1)} \text{Diag} \left( S_\zeta^{-1}(m_1 - m_0) \right) \frac{1}{2}(m_1 - m_0) \quad (8)$$

$$S_\eta = \text{Diag} \left( S_\zeta^{-1}(m_1 - m_0) \right) S_\zeta \text{Diag} \left( S_\zeta^{-1}(m_1 - m_0) \right). \quad (9)$$

With (7), we rewrite the cLLR  $L(k)$  at time  $k$  as the separable sum of  $\eta_i(k)$ 's:

$$L(k) = \sum_{i=1}^N \eta_i(k).$$

### C. Distributed detector: Consensus+Innovations

We now consider the consensus+innovations distributed detector, see (1), with structure like the structure of the distributed estimator in [6] or of the  $\mathcal{MD}$  distributed detector in [5]. The key to ours is our choice of the consensus weight sequence  $\gamma_k^1$  that we show to be the optimal one and will lead to exponential decay rate of the probability of error of the consensus+innovations distributed detector, which is not the case in general for the  $\mathcal{MD}$  distributed detector in [5] as we show in Appendix B.

To set-up the distributed detector, let the decision variable or the current state of sensor  $i$  at time  $k$  be  $x_i(k)$ . Due to the *communication* noise, when sensor  $j$  transmits to sensor  $i$  its state, sensor  $i$  receives a noisy version:

$$x_j(k) + \nu_{ij}(k), \quad (10)$$

where  $\nu_{ij}(k)$  is the *communications* noise. Note that (10) is a high level model, i.e., we do not model here the physical communication channel, but rather we model the estimation errors at the receiver.

We propose as distributed detector the single time scale, stochastic approximation, consensus+innovations algorithm where each sensor updates its decision variable two-fold: 1) by *consensus*, i.e., averaging its decision variable with the decision variables of its immediate neighbors—the sensors with which it communicates; and 2) by *innovation*, i.e., by incorporating the innovation  $\eta_i(k)$  in (7), after *sensing* its local observation. The consensus+innovation update of  $x_i(k)$  is given by:

$$x_i(k+1) = x_i(k) + \frac{k}{k+1} \alpha_k \underbrace{\sum_{j \in O_i} ((x_j(k) - x_i(k)) + \nu_{ij}(k))}_{\text{consensus}} + \frac{1}{k+1} \underbrace{(\eta_i(k+1) - x_i(k))}_{\text{innovations}} \quad (11)$$

$$= \frac{k}{k+1} (1 - \alpha_k d_i) x_i(k) + \frac{k}{k+1} \alpha_k \sum_{j \in O_i} (x_j(k) + \nu_{ij}(k)) + \frac{1}{k+1} \eta_i(k+1) \quad (12)$$

$$x_i(1) = \eta_i(1).$$

We write (12) in matrix form. The *communication* noise at sensor  $i$  from all its neighbors, and the corresponding vector quantity for all sensors, are (see (12))

$$v_i(k) := \sum_{j \in O_i} \nu_{ij}(k), \quad v(k) := (v_1(k), \dots, v_N(k))^T. \quad (13)$$



Denote also by  $S_v := \text{Cov}(v(k))$ . Likewise, define the vector of the sensors decision variables or vector of sensor states  $x(k) := (x_1(k), \dots, x_N(k))^\top$  and the time varying, deterministic, averaging matrix  $W(k) := I - \alpha_k \mathcal{L}$ , where  $\mathcal{L}$  is the graph Laplacian, see Subsection II-A. Then (12) is:

$$x(k+1) = \frac{k}{k+1}W(k)x(k) + \frac{k}{k+1}\alpha_k v(k) + \frac{1}{k+1}\eta(k+1), \quad k = 1, 2, \dots, \quad x(1) = \eta(1), \quad (14)$$

where  $\eta(k)$  is given in (6). When the noises are Gaussian, since (12) and (14) are linear, the decision variables  $x_i(k)$  and  $x(k)$  are Gaussian. For the vector  $x(k)$  of the decision variables  $x_i(k)$ , the vector  $\mu(k)$  of the means  $\mu_i(k)$ ,  $1 \leq i \leq N$  under  $H_1$  (respectively,  $H_0$ ) and the covariance  $S_\mu(k)$  under either hypotheses are:

$$\begin{aligned} \mu(k) &= \mathbb{E}[x(k) | H_1] = -\mathbb{E}[x(k) | H_0] = (\mu_1(k) \ \mu_2(k) \ \dots \ \mu_N(k))^\top \\ S_\mu(k) &= \text{Cov}(x(k)). \end{aligned} \quad (15)$$

We let the diagonal elements of  $S_\mu(k)$  be

$$\sigma_i^2(k) = [S_\mu(k)]_{ii}. \quad (16)$$

**Weight sequences  $\gamma_k^1$  and  $\gamma_k^2$ .** Comparing (11) with (1), the consensus and innovations weights are

$$\gamma_k^1 = \frac{k}{k+1}\alpha_k \quad \text{and} \quad \gamma_k^2 = \frac{1}{k+1}.$$

Due to the *communication* noise, the  $\{\alpha_k\}$  have to be diminishing, i.e.,  $\alpha_k \rightarrow 0$ , as pointed out in [18], [6], [5]. The design of the  $\{\alpha_k\}$  will be key to the distributed detector achieving exponential decay rate of the error probability: a small and fast-decaying  $\alpha_k$  injects low communication noise in the decision variable  $x_i(k)$ , but limits the information flow among neighbors (insufficient averaging). We will show that, for *appropriately designed constants*  $a, b_0 > 0$ ,

$$\alpha_k = \frac{b_0}{a+k},$$

balances these two opposing effects—communication noise and information flow. As detailed in Section IV, a large  $b_0$  yields larger noise injection but also greater inter-sensor averaging.

We compare the consensus+innovations distributed detector (12) with the distributed detectors in [4] and [5]. The detector in [4], referred to as running consensus, uses constant, non-decaying weights  $\alpha_k = \alpha$ , which is not suitable for noisy communication. To account for communication noise, references [6], [5] propose mixed time scale, stochastic approximation type algorithms. In particular, for detection, [5] pro-

poses the  $\mathcal{MD}$  algorithm for the generic case of different signal-to-noise ratios (**SNR**) at different sensors and the single time scale  $\mathcal{SD}$  algorithm when the **SNR** is the same at all sensors. The algorithm  $\mathcal{MD}$  uses the weight sequences (see eqn. (13), [5])

$$\text{Consensus : } \quad \left\{ \frac{c_1}{k^\tau} \right\}, \quad \tau \in (0.5; 1), \quad c_1 > 0 \quad (17)$$

$$\text{Innovations : } \quad \left\{ \frac{c_2}{k} \right\}, \quad c_2 > 0 \quad (18)$$

The algorithm  $\mathcal{SD}$  uses the same weights for both consensus and innovations (see eqn. (53) in [5]) equal to (17) and (18) with  $\tau = 1$ . In contrast with [5], we propose the single time scale algorithm (12) regardless if the local **SNR** are mutually different or not. A major contribution here with respect to [5] is to show that the single time scale algorithm (12) yields better asymptotic detection performance than the mixed time scale algorithm  $\mathcal{MD}$ . Our algorithm (12) and  $\mathcal{SD}$  are both single time scale. The main differences between (12) and  $\mathcal{SD}$  are that algorithm (12): 1) incorporates  $a$  in (II-C); and 2) optimizes the parameter  $b_0$ . We will show that (12) exhibits under appropriate structural conditions exponential rate of decay of the probability of error at every sensor, while  $\mathcal{MD}$  in [5] is sub exponential;  $\mathcal{SD}$  is shown to be exponential, but only when the sensors are identical, i.e., all operate under the same **SNR**.

#### D. Signal-to-noise ratios (**SNR**)

We define, for future reference, the following relevant **SNRs**.

##### 1. Sensing **SNR**:

$$\text{global : } \quad \mathbf{SSNR} := (m_1 - m_0) S_\zeta^{-1} (m_1 - m_0) \quad (19)$$

$$\text{local : } \quad \mathbf{SSNR}_i := \frac{([m_1 - m_0]_i)^2}{[S_\zeta]_{ii}}. \quad (20)$$

##### 2. Detector **SNR**, for a generic detector gen (either centralized, isolated, or distributed):

$$\mathbf{DSNR}_{\text{gen}}(k) := \frac{\mathbb{E}^2 [\mathcal{D}_{\text{gen}}(k) | H_1]}{\text{Var} (\mathcal{D}_{\text{gen}}(k) | H_1)},$$

where  $\mathcal{D}_{\text{gen}}(k)$  is the detector decision variable. We denote the decision **SNR** for the centralized, isolated, and distributed detector, respectively, by  $\mathbf{DSNR}(k)$ ,  $\mathbf{DSNR}_i(k)$ , and  $\mathbf{DSNR}_{\text{dis},i}(k)$ . We will see later how detector **SNR** determines the error probability for Gaussian detectors; see ahead (25).

3. *Communication **SNR*** is a quantity that accounts for the *communication noise* and plays a role only with the distributed detector. We define the communication **SNR** (per sensor) by:

$$\mathbf{CSNR} = \frac{(\frac{1}{N} \mathbf{SSNR})^2}{\|S_v\|}, \quad (21)$$

where we recall the *communication* noise covariance  $S_v = \text{Cov}(v(k))$ .

**Remark.** We give a hint why **CSNR** is defined as (21) and why it plays a significant role in assessing distributed detection performance. With our distributed detector, sensors communicate their local decision variables  $x_i(k)$ ;  $x_i(k)$  is a local approximation of the (scaled) centralized decision variable  $\frac{1}{N}\mathcal{D}(k)$  (see (5)). The mean of  $x_i(k)$  under  $H_1$ , for large  $k$ , is close to the mean of  $\frac{1}{N}\mathcal{D}(k)$ , equal to  $\frac{1}{2N}\mathbf{SSNR}$  (as will be shown); the variance of  $x_i(k)$ , as will be shown, vanishes at rate  $1/k$ . Hence, **CSNR** describes how well, in a sense, the signal  $x_i(k)$  competes against the noise  $v_i(k)$  in communication, for large  $k$ . We define also the communication gain as the ratio of the communication SNR and the average (across sensors) sensing SNR<sup>4</sup>:

$$\mathbf{G}_c = \frac{\mathbf{CSNR}}{\frac{1}{N}\mathbf{SSNR}} = \frac{\frac{1}{N}\mathbf{SSNR}}{\|S_v\|}. \quad (22)$$

For future reference, we introduce here the following two constants that we will need when assessing distributed detection performance:

$$c_\mu := \frac{2\sqrt{N}\|m_\eta^{(1)}\|}{\frac{1}{N}\mathbf{SSNR}}, \quad c_\sigma := \frac{\|S_\eta\|}{\frac{1}{N}\mathbf{SSNR}}. \quad (23)$$

### III. MODELING ASSUMPTIONS AND PRELIMINARY RESULTS

In this Section, we establish our underlying assumptions and present the asymptotic performance of the isolated sensor and centralized detectors. These will be a prelude to our main results on the asymptotic performance of the consensus+innovations detector in (12) given in Section IV and proven in Section V and Appendix A.

#### A. Modeling assumptions

As mentioned in Section II, the noises are zero mean Gaussian spatially correlated but temporally independent sequences. In Section VI, we will consider the case where the noises are not Gaussian.

*Assumption 1 (Gaussian noises)* The *sensing* and the *communication* noises  $\zeta_i(k)$  and  $v_{ij}(k)$  are zero mean, Gaussian spatially correlated and temporally independent noises, and independent of each other:

$$\zeta(k) \sim \mathcal{N}(0, S_\zeta), \quad v(k) \sim \mathcal{N}(0, S_v),$$

where the vector  $v(k)$  is defined in (13) and  $S_\zeta$  and  $S_v$  are assumed to be positive definite.

<sup>4</sup>Note that **CSNR** and **SSNR** are not independent quantities here; larger **SSNR** means larger **CSNR**.

In distributed processing, the ability for the sensors or agents to cooperate is fundamental; this is captured by the connectivity of the network.

*Assumption 2 (Network connectedness)* The network  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  is connected.

As it is well known, a necessary and sufficient condition for connectedness is  $\lambda_2(\mathcal{L}) > 0$ , i.e., the algebraic connectivity of the network is strictly positive. The next assumption is on the weight sequence  $\{\alpha_k\}$ .

*Assumption 3 (Weight sequence)* The weight sequence  $\{\alpha_k\}$  is:

$$\alpha_k = \frac{a}{b_0 + k},$$

where the constants  $a$  and  $b_0$  satisfy

$$a \geq b_0 \lambda_N(\mathcal{L}) > 0, \quad b_0 > \max \left\{ 0, \frac{c_\mu - 1}{\lambda_2(\mathcal{L})} \right\}. \quad (24)$$

The role of these conditions will become clear when we state our main result, Theorem 3, in Section IV. Recall the sensing **SNRs** in (19). We make the following assumption on **SSNR**.

*Assumption 4* **SSNR**  $> 0$ .

Note that Assumption 4 is equivalent to having different mean vectors,  $m_1 \neq m_0$ . To obtain certain specialized results, we will assume a stronger assumption than Assumption 4.

*Assumption 5 (Equal local sensing **SNRs**)* **SSNR** $_i = \mathbf{SSNR}_j > 0, \quad \forall i \neq j$ .

### B. Asymptotic performance and Chernoff information

For Gaussian decision variables like for the three detectors (isolated sensor, centralized, and consensus+innovations distributed), and equal prior probabilities, the probability of error  $P^e(k)$  is given by

$$P^e(k) = \mathcal{Q} \left( \sqrt{\mathbf{DSNR}_{\text{gen}}(k)} \right). \quad (25)$$

where  $\mathcal{Q}$  is the  $Q$ -function in (2) and  $\mathbf{DSNR}_{\text{gen}}(k)$  is the generic detector **SNR**.

To determine the exponential decay rate of the error probability, we recall the bounds [23]

$$\frac{t}{1+t^2} e^{-t^2/2} \leq 2\pi \mathcal{Q}(t) \leq \frac{1}{t} e^{-t^2/2}, \quad t > 0, \quad (26)$$

We apply these bounds to (25). Taking the logarithm and dividing by  $k$ , the limsup of the right hand side (rhs) inequality and the liminf of the left hand side (lhs) inequality in (26) lead to <sup>5</sup>:

$$\limsup_{k \rightarrow \infty} -\frac{1}{k} \log P^e(k) \leq \limsup_{k \rightarrow \infty} \frac{1}{2k} \mathbf{DSNR}_{\text{gen}}(k) \quad (27)$$

$$\liminf_{k \rightarrow \infty} -\frac{1}{k} \log P^e(k) \geq \liminf_{k \rightarrow \infty} \frac{1}{2k} \mathbf{DSNR}_{\text{gen}}(k). \quad (28)$$

If the limsup in (27) is zero, we have two possibilities: 1) the error probability  $P^e(k)$  decays to zero slower than exponentially in  $k$ ; or 2)  $P^e(k)$  does not converge to zero at all. Intuitively, large mean and fast-shrinking variance of the decision variables  $\mathcal{D}_{\text{gen}}(k)$  increase  $\mathbf{DSNR}_{\text{gen}}(k)$  and hence yield good detection performance.

The detector for the isolated sensor  $i$  given by equation (4) and the centralized detector in (5) are the optimal minimum probability of error detectors (when prior probabilities are equal, the threshold  $\tau(k) = 0$ .) For these detectors, the limsup and the liminf in (27) and (28) actually coincide, i.e., the sequence  $\frac{1}{2k} \mathbf{DSNR}_{\text{gen}}(k)$  has a limit and

$$C_{\text{gen}} = \lim_{k \rightarrow \infty} -\frac{1}{k} \log P_{\text{gen}}^e(k) = \lim_{k \rightarrow \infty} \frac{\mathbf{DSNR}_{\text{gen}}(k)}{2k}. \quad (29)$$

These detectors maximize the exponential decay rate of the probability of error for the corresponding problems; this optimal exponential decay rate is the Chernoff information  $C_{\text{gen}}$  as indicated in (29). For the isolated sensor  $i$  detector and the centralized detector, it is easily shown that their detectors  $\mathbf{SNR}_{\text{DSNR}_i}(k)$  and  $\mathbf{DSNR}(k)$  are given by<sup>6</sup>

$$\mathbf{DSNR}_i(k) = \frac{k}{4} \mathbf{SSNR}_i, \quad \mathbf{DSNR}(k) = \frac{k}{4} \mathbf{SSNR}. \quad (30)$$

The Chernoff information for the isolated and the centralized optimal detectors are then:

$$C_i = \lim_{k \rightarrow \infty} -\frac{1}{k} \log P_i^e(k) = \lim_{k \rightarrow \infty} \frac{\mathbf{DSNR}_i(k)}{2k} = \frac{1}{8} \mathbf{SSNR}_i \quad (31)$$

$$C = \lim_{k \rightarrow \infty} -\frac{1}{k} \log P_{\text{cen}}^e(k) = \lim_{k \rightarrow \infty} \frac{\mathbf{DSNR}(k)}{2k} = \frac{1}{8} \mathbf{SSNR}. \quad (32)$$

Eqns. (31) and (32) justify the following definition of the global and local detectability, after which we relate global and local detectability to sensing (global and local)  $\mathbf{SNRs}$  by a simple, but important fact.

<sup>5</sup>Eqn. (28) holds because  $\mathbf{DSNR}_{\text{gen}}(k)$  is strictly positive (for large  $k$ ) and can grow at most as  $\sim k$  for either centralized, isolated, or distributed detectors (as will be shown).

<sup>6</sup>We will see later that  $\mathbf{DSNR}_{i,\text{dis}}(k)$  grows also at rate  $k$  with our distributed detector (12).

*Definition 1* The network  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  is globally detectable if the probability of error  $P^e(k)$  of the optimal centralized detector decays exponentially fast. The sensor  $i \in \mathcal{V}$  is locally detectable if the probability of error  $P_i^e(k)$  of its optimal detector in isolation decays exponentially fast.

*Fact 2* The network  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  is globally detectable if and only if  $\mathbf{SSNR} > 0$ . The sensor  $i \in \mathcal{V}$  is locally detectable if and only if  $\mathbf{SSNR}_i > 0$ .

Clearly, a network can be globally detectable but have many (or most) sensors that are not locally detectable; however, at least one sensor needs to be locally detectable so that global detectability holds. Our goal is to carry out a similar asymptotic performance analysis for the distributed consensus+innovations detector in (12) or in vector form in (14). Because  $x_i(k)$  is Gaussian, relations (28) and (27) still apply. Recall the mean and variance of the decision variable  $x_i(k)$  under  $H_1$ ,  $\mu_i(k)$  and  $\sigma_i^2(k)$ , given by (15) and (16). The distributed detector  $\mathbf{SNR}$ ,  $\mathbf{DSNR}_{\text{dis},i}(k)$ , at sensor  $i$  at time  $k$  is then:

$$\mathbf{DSNR}_{\text{dis},i}(k) = \frac{(\mu_i(k))^2}{\sigma_i^2(k)}.$$

We obtain the moments  $\mu_i(k)$  and  $\sigma_i^2(k)$  of  $x_i(k)$  and their asymptotic values in the next Section IV by analyzing the distributed algorithm in (12). In contrast with the centralized and isolated detectors, these statistics of the decision variable  $x_i(k)$  of the  $i$ th sensor are affected by the communication noise  $\nu_{ij}(k)$ , see equation (10), through  $v_i(k)$  and  $v(k)$  in (13) and (13). We will see that, besides  $\mathbf{DSNR}_{\text{dis},i}(k)$ , we need to account for the impact of the  $\mathbf{G}_c$  given in (22).

#### IV. CONSENSUS+INNOVATIONS DISTRIBUTED DETECTION: PERFORMANCE ANALYSIS

Subsection IV-A studies the exponential decay of our consensus+innovations detector in (12), subsection IV-B addresses the optimality of the weight sequence  $\{\alpha_k\}$ , and subsection IV-C addresses the potential payoff of distributed detection arising from noisy cooperation among sensors.

##### A. Exponential decay of the error probability

The next Theorem establishes under reasonable conditions that the probability of error at every sensor of the consensus+innovations distributed detector in (12) decays exponentially fast. Recall the definitions of  $\mathbf{SSNR}$  and  $\mathbf{G}_c$  in (19) and (22), and the constants  $c_\mu$  and  $c_\sigma$  in (23).

*Theorem 3* Consider the consensus+innovations distributed detector in (12) under the Assumptions 1, 2, 3, and 4. Then:

1) The moments  $\mu(k)$ ,  $\mu_i(k)$ , and  $\sigma_i^2(k)$  satisfy:

$$\mu_\infty := \lim_{k \rightarrow \infty} \mu(k) = (I + b_0 \mathcal{L})^{-1} m_\eta^{(1)} \quad (33)$$

$$\lim_{k \rightarrow \infty} \mu_i(k) \geq \frac{1}{2N} \text{SSNR} \left( 1 - \frac{\sqrt{N}}{1 + b_0 \lambda_2(\mathcal{L})} c_\mu \right) \quad (34)$$

$$\limsup_{k \rightarrow \infty} k \sigma_i^2(k) \leq \frac{1}{N^2} \text{SSNR} \left( 1 + 3 \frac{N}{1 + b_0 \lambda_2(\mathcal{L})} c_\sigma + \frac{N b_0^2}{\mathbf{G}_c} \right). \quad (35)$$

2) The exponential decay rate of the error probability  $P_{\text{dis},i}^e(k)$  at every sensor  $i$  satisfies:

$$\liminf_{k \rightarrow \infty} -\frac{1}{k} \log P_{\text{dis},i}^e(k) \geq \frac{1}{8} \text{SSNR} \frac{\left( 1 - \frac{\sqrt{N}}{1 + b_0 \lambda_2(\mathcal{L})} c_\mu \right)^2}{1 + 3 \frac{N}{1 + b_0 \lambda_2(\mathcal{L})} c_\sigma + \frac{N b_0^2}{\mathbf{G}_c}}. \quad (36)$$

Before proving the Theorem, which we carry out in Section V, we analyze how the bound on the rhs of (36) depends on the different **SNRs**, on the network connectivity  $\lambda_2(\mathcal{L})$ , and on the parameter  $b_0$  of the weight sequence  $\{\alpha_k\}$ . The discussion is summarized in the following five remarks on Theorem 3.

**1. Exponential decay of the error probability  $P_{\text{dis},i}^e$ .** Under global detectability and connectedness, Theorem 3 states that the error probability  $P_{\text{dis},i}^e$  at *every* sensor  $i$  decays exponentially to zero *even if sensor  $i$  is (in isolation) not detectable* ( $\text{SSNR}_i = 0$ ), and even when the communication links are very noisy ( $\mathbf{G}_c > 0$  but small.) This feature of the distributed detector (12) significantly improves over existing work like  $\mathcal{MD}$  in [5]. Namely, we prove in Appendix B that  $\mathcal{MD}$  achieves only a sub exponential decay rate of order  $e^{-c k^\tau}$ ,  $\tau < 1$ ,  $c > 0$ , of the error probability, irrespective of  $\mathbf{G}_c$ .

**2. Effect of  $\mathbf{G}_c$ .** The bound on the rhs of (36) shows quantitatively that higher  $\mathbf{G}_c$  leads to better detection, confirming the qualitative discussion in the Remark below (21).

**3. Effect of the network connectivity  $\lambda_2(\mathcal{L})$ .** Theorem 3 shows that the network connectivity plays a role in the detection performance through the algebraic connectivity  $\lambda_2(\mathcal{L})$ . Larger values of  $\lambda_2(\mathcal{L})$ , which allow for faster averaging, increase the bound (36) yielding faster decay rate for the error probability.

**4. Tradeoff: Communication noise vs. information flow.** With optimal centralized detection (that corresponds to a fully connected network and no additive communication noise) we have that, for all  $i$ ,  $P_i^e(k) \equiv P^e(k)$ , and:  $\lim_{k \rightarrow \infty} -\frac{1}{k} \log P_i^e(k) = \frac{1}{8} \text{SSNR}$ . Then, from (36), all the three terms:

$$\frac{\sqrt{N}}{1 + b_0 \lambda_2(\mathcal{L})} c_\mu \quad (37)$$

$$3 \frac{N}{1 + b_0 \lambda_2(\mathcal{L})} c_\sigma \quad (38)$$

$$\frac{N b_0^2}{\mathbf{G}_c} = \frac{N}{b_0^2} \quad (39)$$

decrease the bound and so they quantify the decrease in performance of the distributed detector with respect to the centralized detector. This decrease comes from two effects: 1) communication noise; and 2) insufficient information flow.

From (37)–(39), we can see how the parameter  $b_0$  affects in opposing ways these two effects: The terms (37) and (38) relate to the information flow, while the term (39) is due to communication noise. We see that the net effect of increasing  $b_0$  is to increase the effective algebraic connectivity ( $b_0$  multiplies  $\lambda_2(\mathcal{L})$ ), increasing (37) and (38); on the other hand, it reduces **CSNR** as seen from (39).

The weight choice  $\alpha_k$  in (II-C) optimally balances these two effects if we tune the parameter  $b_0$  to maximize the right hand side in (36). This is a scalar optimization problem in  $b_0$  and can be easily numerically performed. Lemma 6 find the optimal  $b_0$  in closed form for a simplified case.

**5. Tradeoff: Bias-variance.** Theorem 3 reveals a certain bias-variance tradeoff. Ideally, we would like the bias-free decision variables:

$$\mu_\infty = \left( \frac{1}{2N} \mathbf{SSNR} \right) \mathbf{1}, \quad (40)$$

where  $\mathbf{1}$  is the vector of ones; i.e., all sensors should have as asymptotic decision variable the asymptotic centralized decision variable. That is, we want the mean of the decision variable at each sensor to converge to the expected value of the centralized decision variable  $\mathcal{D}(k)$  (divided by  $1/N$ .) Our algorithm (12) introduces a bias (see (33) and (34)), but, on the other hand, it decreases the variance at the optimal rate  $1/k$ . In contrast,  $\mathcal{MD}$  in [5] does not have the bias, but it decreases the variance at a slower rate. Compared to  $\mathcal{MD}$ , our algorithm (12) better resolves the bias-variance tradeoff in terms of the detection performance; algorithm (12) decays the error probability exponentially, while  $\mathcal{MD}$  decays it sub exponentially. We now consider a special case where all sensors are identical, or, better said, they operate under the same  $\mathbf{SSNR}_i$ , i.e., Assumption 5 holds. Theorem 3 takes a simplified form, where  $\mu_\infty$  becomes bias free, as in (40). Further,  $c_\mu = c_\sigma = 1$  and second condition in (24) becomes  $b_0 > 0$ ; it can also be shown (details omitted) that the factor 3 in (38) reduces to 1. The simplified Theorem 3 follows.

*Theorem 4 (Asymptotic performance: Identical sensors)* Let Assumptions 1 through 3 and 5 hold. Then, the exponential decay rate of the error probability at each sensor  $i$  satisfies:

$$\liminf_{k \rightarrow \infty} -\frac{1}{k} \log P_{\text{dis},i}^e(k) \geq \frac{1}{8} \mathbf{SSNR} \frac{1}{\left( 1 + \frac{N}{1+b_0\lambda_2(\mathcal{L})} + \frac{Nb_0^2}{\mathbf{G}_e} \right)} \quad (41)$$

$$\geq \frac{1}{8} \mathbf{SSNR} \frac{1}{\left( 1 + \frac{N}{b_0\lambda_2(\mathcal{L})} + \frac{Nb_0^2}{\mathbf{G}_e} \right)} \quad (42)$$



### B. Optimality of the weight sequence $\{\alpha_k\}$

**Order-optimality.** We consider the role of the weight sequence (II-C), in particular, we show the optimality of the rate  $1/k$ . To this end, we consider the distributed detector (12) but modify the weight sequence; we refer to the modified sequence as  $\beta_k$ . We find an upper bound on the decay rate of the error probability when the weight sequence is re-set to  $\alpha_k$ .

*Theorem 5* Let Assumptions 1—3 and 5 hold. Suppose that the weight choice  $\alpha_k$  in (II-C) is replaced by:

$$\beta_k = \frac{b_0}{(a + k^\tau)}, \quad a, b_0 > 0,$$

where  $\tau \geq 0$ . We have:

$$\limsup_{k \rightarrow \infty} -\frac{1}{k} \log P_{\text{dis},i}^e(k) \leq \begin{cases} 0 & \text{if } \tau < 1 \\ \frac{1}{8} \text{SSNR}_i & \text{if } \tau > 1 \\ \frac{1}{8} \text{SSNR} \frac{1}{1 + \frac{N}{1+2b_0\lambda_N(\mathcal{L})} + \frac{b_0^2 \lambda_1(S_v)}{N \text{SSNR}_i}} & \text{if } \tau = 1. \end{cases} \quad (43)$$

*Proof:* See Appendix A. ■

Two remarks on Theorem 5 are in order.

**1. Order-optimality of  $\alpha_k$ .** Theorem 5 says that the choice  $\alpha_k$  in (II-C) is the optimal weight choice in the family of choices  $\beta_k = \frac{b_0}{(a+k^\tau)}$ ,  $a, b_0 > 0$ , parametrized by  $\tau \geq 0$ . If  $\beta_k$  decays too slowly ( $\tau < 1$ ), then the error probability converges to zero at a rate slower than exponential (if at all it converges to zero.) On the other hand, if  $\beta_k$  decays too fast ( $\tau > 1$ ), then the error probability does decay to zero exponentially, but the rate is no better than the rate of the individual detection, irrespective of  $\mathbf{G}_c$ .

**2. Tightness of the bounds in Theorems 4 and 5.** The upper bound (43) for  $\tau = 1$  explains the tightness of the lower bound in Theorem 4 and the unavoidable simultaneous effects of the communication noise and information flow. The sequence  $\{\alpha_k\}$  balances these via the parameter  $b_0$ .

**Optimal  $b_0^*$ .** We now find  $b_0^*$  that optimizes (maximizes) (42); we pursue (42) rather than (41) as it allows simpler, closed form expressions. Proof of Lemma 6 follows after setting the derivative of the denominator of rhs in (42) to zero and is hence omitted for brevity.

*Lemma 6* Let Assumptions 1 through 3 and 5 hold. The optimal parameter  $b_0^*$  that maximizes (42) and

the corresponding lower bound on  $\liminf_{k \rightarrow \infty} -\frac{1}{k} \log P_i^e(k)$ , are, respectively:

$$b_0^* = \frac{\mathbf{G}_c^{1/3}}{\lambda_2(\mathcal{L})^{1/3} 4^{1/3}} \quad (44)$$

$$\liminf_{k \rightarrow \infty} -\frac{1}{k} \log P_{\text{dis},i}^e(k) \geq \frac{1}{8} \mathbf{SSNR} \frac{1}{1 + \frac{N c_0}{\lambda_2(\mathcal{L})^{2/3} \mathbf{G}_c^{1/3}}}, \quad (45)$$

where

$$c_0 = \frac{1}{2} 4^{1/3} + 4^{-2/3} = \frac{3}{2} (2)^{-1/3} \approx 1.19.$$

We use Lemma 6 to compare the distributed detector with the optimal centralized detector and the optimal single sensor detector. In the very high  $\mathbf{G}_c$  regime (weak communication noise), when  $\mathbf{G}_c \rightarrow \infty$ , the distributed detector (at all sensors) achieves the asymptotic performance of the optimal centralized detector. On the other hand, when  $\mathbf{G}_c$  decreases, at some point, the rhs in (45) falls below  $\frac{1}{8} \mathbf{SSNR}_i$  and the distributed detector (12) at sensor  $i$  becomes worse than if sensor  $i$  worked in isolation. The discussion is formalized in the next Subsection that considers when sensors should cooperate.

### C. Communication payoff

Eqn. (45) under low  $b f G_c$  raises the issue whether a sensor  $i$  should cooperate with its neighbors or not. We next formalize communication payoff.

*Definition 7 (Communication payoff)* The network  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  achieves communication payoff if:

$$\min_{i=1,\dots,N} \left\{ \liminf_{k \rightarrow \infty} -\frac{1}{k} \log P_{\text{dis},i}^e(k) \right\} \geq \max_{i=1,\dots,N} \left\{ \frac{1}{8} \mathbf{SSNR}_i \right\}.$$

Definition 7 says that the network achieves a communication payoff if the distributed detector error performance of the worst sensor is better than the isolated detector error performance for the best sensor without communication. Lemma 8 finds a threshold on the  $\mathbf{G}_c$  above which it does pay off for sensors to communicate with their neighbors. Proofs of Lemma 8 is simple and is omitted.

*Lemma 8* Let Assumptions 1 through 3 and 5 hold. Set  $b_0$  to the optimal value in (44). If

$$\mathbf{G}_c \geq \left( c_0 \frac{N}{N-1} \right)^3 \left( \frac{1}{\lambda_2(\mathcal{L})} \right)^2,$$

then the network achieves the communication payoff in the sense of Definition 7.

## V. PROOF OF THEOREM 3

Subsection V-A sets up the analysis and Subsection V-B proves Theorem 3.

### A. Solution of the consensus+innovations distributed detector

Define the matrices  $\Phi(k, j)$ ,  $k \geq j \geq 1$ , as follows:

$$\Phi(k, j) := \begin{cases} W(k-1)W(k-2)\dots W(j) & \text{if } 1 \leq j < k \\ I & \text{if } j = k. \end{cases}$$

Then, the solution to the distributed detector (14) is:

$$x(k) = \frac{1}{k} \sum_{j=1}^k \Phi(k, j) \eta(j) + \frac{1}{k} \sum_{j=1}^{k-1} (j \alpha_j) \Phi(k, j+1) v(j), \quad k = 1, 2, 3, \dots \quad (46)$$

Introduce

$$\begin{aligned} \widetilde{W}(k) &:= W(k) - J \\ \widetilde{\Phi}(k, j) &:= \widetilde{W}(k-1)\widetilde{W}(k-2)\dots\widetilde{W}(j), \quad k > j \geq 1. \end{aligned}$$

It can be seen that

$$\widetilde{\Phi}(k, j) = \Phi(k, j) - J.$$

In consensus,  $W(k) \rightarrow J$ , where  $J$  is the ideal consensus averaging matrix. The matrix  $\widetilde{W}(k)$  and its norm  $\|\widetilde{W}(k)\|$  measure, in a sense, the imperfection in the information flow, i.e., how far  $\widetilde{W}(k)$  is away from 0 and  $W(k)$  from  $J$ . If (24) holds then it is easy to see that

$$\|\widetilde{W}(k)\| = 1 - \frac{b_0 \lambda_2(\mathcal{L})}{a + k} \in \left(1 - \frac{\lambda_2(\mathcal{L})}{\lambda_N(\mathcal{L})}, 1\right), \quad \forall k. \quad (47)$$

We see that the role of  $a$  in  $\alpha_k$  is to be an offset that enables (47) to hold for all  $k$ ; that is,  $a$  reduces  $\|\widetilde{W}(k)\|$  for large  $b_0$  and small  $k$ . We will see that  $b_0$  is the effective tuning parameter that controls the detection performance. We also comment that the ratio  $\frac{\lambda_2(\mathcal{L})}{\lambda_N(\mathcal{L})}$  is maximized by Ramanujan networks, see [24] for details.

### B. Proof

*Proof of Theorem 3: Claims (33) and (34):* We first study the mean of the decision variable  $\mu(k)$ . It evolves according to the following recursion (which can be seen by taking the expectation in (14)):

$$\mu(k+1) = \frac{k}{k+1} W(k) \mu(k) + \frac{m_\eta^{(1)}}{k+1}. \quad (48)$$

Next, we consider the *error*  $\epsilon(k)$  of  $\mu(k)$  wrt the assumed  $\mu_\infty$  given in (33):

$$\epsilon(k) := \mu(k) - (I + b_0 \mathcal{L})^{-1} m_\eta^{(1)}.$$

We will show that  $\epsilon(k) \rightarrow 0$ , which implies (33). Algebraic manipulations show that  $\epsilon(k)$  satisfies:

$$\epsilon(k+1) = \frac{k}{k+1} W(k) \epsilon(k) + \frac{1}{k+1} \Gamma(k) m_\eta^{(1)}, \quad (49)$$

where

$$\begin{aligned} \Gamma(k) &= I - (k+1) (I + b_0 \mathcal{L})^{-1} + k W(k) (I + b_0 \mathcal{L})^{-1} \\ &= I - (k+1) (I + b_0 \mathcal{L})^{-1} + k \left( I - \frac{b_0}{a+k} \mathcal{L} \right) (I + b_0 \mathcal{L})^{-1}. \end{aligned}$$

Recall the eigendecomposition of the Laplacian in (3). The matrix  $\Gamma(k)$  has the same eigenvectors as  $\mathcal{L}$ ; simple calculations show that the eigenvalue  $\lambda_i(\Gamma(k))$  that corresponds to the eigenvector  $q_i$  is:

$$\lambda_i(\Gamma(k)) = \begin{cases} 0 & \text{if } i = 1 \\ \frac{b_0 a \lambda_i(\mathcal{L})}{1 + b_0 \lambda_i(\mathcal{L})} \frac{1}{k+a} =: \frac{c_{\Gamma,i}}{k+a} & \text{otherwise.} \end{cases}$$

Then, clearly, for some  $c_\Gamma > 0$ ,

$$\|\Gamma(k)\| \leq \frac{c_\Gamma}{k}. \quad (50)$$

We now decompose  $\epsilon(k)$  into the *consensus subspace*, i.e., the component colinear with the vector  $\mathbf{1}$ , and the component orthogonal to  $\mathbf{1}$ :  $\epsilon(k) = (I - J) \epsilon(k) + J \epsilon(k) = (I - J) \epsilon(k) + \left( \frac{1}{N} \mathbf{1}^\top \epsilon(k) \right) \mathbf{1}$ . We show separately that:

$$\lim_{k \rightarrow \infty} (I - J) \epsilon(k) = 0 \quad (51)$$

$$\lim_{k \rightarrow \infty} \mathbf{1}^\top \epsilon(k) = 0. \quad (52)$$

Then, (51) and (52) together imply that

$$\lim_{k \rightarrow \infty} \epsilon(k) = 0. \quad (53)$$

We first show (52). Multiplying (49) from the left by  $\mathbf{1}^\top$ , using the orthogonality of the eigenvectors  $q_i$ , and using the fact that  $\mathbf{1}^\top W(k) = \mathbf{1}^\top$ , we get:

$$\mathbf{1}^\top \epsilon(k+1) = \frac{k}{k+1} \mathbf{1}^\top \epsilon(k),$$

which implies (52).

We now show (51). Denote by

$$b := b_0 \lambda_2(\mathcal{L}).$$

Multiplying (49) from the left by  $(I - J)$ , we get:

$$\begin{aligned} (I - J) \epsilon(k + 1) &= \frac{k}{k + 1} (I - J) W(k) \epsilon(k) + \frac{1}{k + 1} (I - J) \Gamma(k) m_\eta^{(1)} \\ &= \frac{k}{k + 1} \widetilde{W}(k) (I - J) \epsilon(k) + \frac{1}{k + 1} \Gamma(k) m_\eta^{(1)}, \end{aligned} \quad (54)$$

where (54) holds because  $JW(k) = J$ ,  $\widetilde{W}(k)J = 0$ , and  $(I - J)\Gamma(k) = \Gamma(k)$ . Now, by subadditivity and submultiplicativity of norms, (54) yields:

$$\begin{aligned} \|(I - J) \epsilon(k + 1)\| &\leq \frac{k}{k + 1} \|\widetilde{W}(k)\| \|(I - J) \epsilon(k)\| + \frac{1}{k + 1} \|\Gamma(k)\| \|m_\eta^{(1)}\| \\ &\leq \left(1 - \frac{b_0 \lambda_2(\mathcal{L})}{a + k}\right) \|(I - J) \epsilon(k)\| + \frac{c_\Gamma \|m_\eta^{(1)}\|}{k^2} \\ &= \left(1 - \frac{b}{a + k}\right) \|(I - J) \epsilon(k)\| + \frac{c_\epsilon}{k^2} \\ &= \|(I - J) \epsilon(k)\| - \frac{b}{a + k} \|(I - J) \epsilon(k)\| + \frac{c_\epsilon}{k^2}. \end{aligned} \quad (55)$$

Before proceeding, we invoke the following deterministic variant of a result due to Robbins and Siegmund (Lemma 11, Chapter 2.2., [25].)

*Lemma 9 ([25])* Let  $\{u(k)\}$ ,  $\{\rho(k)\}$ , and  $\{\kappa(k)\}$  be non-negative deterministic (scalar) sequences. Further, suppose that

$$u(k + 1) \leq u(k) - \rho(k) + \kappa(k), \quad k = 1, 2, \dots$$

Suppose that  $\sum_{k=1}^{\infty} \kappa(k) < \infty$ . Then: 1)  $\sum_{k=1}^{\infty} \rho(k) < \infty$ ; and 2)  $\lim_{k \rightarrow \infty} u(k) = u^*$  exists.

We apply Lemma 9 to (55) with

$$u(k) = \|(I - J) \epsilon(k)\|, \quad \rho(k) = \frac{b}{a + k} \|(I - J) \epsilon(k)\|, \quad \kappa(k) = \frac{c_\epsilon}{k^2}.$$

This proves that

$$\lim_{k \rightarrow \infty} \|(I - J) \epsilon(k)\| = 0,$$

i.e., proves (51). Namely, by Lemma 9, we have that

$$\sum_{k=1}^{\infty} \rho(k) = \sum_{k=1}^{\infty} \frac{b}{a + k} \|(I - J) \epsilon(k)\| < \infty,$$

which implies that

$$\liminf_{k \rightarrow \infty} \|(I - J)\epsilon(k)\| = 0.$$

Also, by Lemma 9,  $\lim_{k \rightarrow \infty} u(k) = \lim_{k \rightarrow \infty} \|(I - J)\epsilon(k)\|$  exists, and, hence,  $\lim_{k \rightarrow \infty} \|(I - J)\epsilon(k)\| = 0$ .

This completes the proof of (53).

We now prove (34) using (33). Note first that

$$(I + b_0 \mathcal{L})^{-1} = Q \Lambda ((I + b_0 \mathcal{L})^{-1}) Q^\top, \quad (56)$$

where

$$\Lambda((I + b_0 \mathcal{L})^{-1}) = \text{Diag} \left( 1, (1 + b_0 \lambda_2(\mathcal{L}))^{-1}, \dots, (1 + b_0 \lambda_N(\mathcal{L}))^{-1} \right).$$

Thus, using the fact that  $q_1 = \frac{1}{\sqrt{N}} \mathbf{1}$  and  $J = q_1 q_1^\top$ , the matrix  $(I + b_0 \mathcal{L})^{-1}$  decomposes as:

$$(I + b_0 \mathcal{L})^{-1} = J + Q \Lambda' Q^\top, \quad (57)$$

with  $\Lambda' = \text{Diag} \left( 0, (1 + b_0 \lambda_2(\mathcal{L}))^{-1}, \dots, (1 + b_0 \lambda_N(\mathcal{L}))^{-1} \right)$ . Multiplying (57) from the right by  $m_\eta^{(1)}$ , and using (8), we get that the entry  $[\mu_\infty]_i$  equals

$$[\mu_\infty]_i = \frac{1}{2N} \mathbf{SSNR} + \left[ Q \Lambda' Q^\top m_\eta^{(1)} \right]_i, \quad (58)$$

Finally, the inequality  $\left[ \left[ Q \Lambda' Q^\top m_\eta^{(1)} \right]_i \right] \leq \|\Lambda'\| \|m_\eta^{(1)}\| = \frac{1}{1 + b_0 \lambda_2(\mathcal{L})} \|m_\eta^{(1)}\|$  yields (34).  $\blacksquare$

We now prove (35); we use the following auxiliary result.

*Lemma 10* Denote by:

$$\begin{aligned} \mathcal{Z}(k) &:= \frac{1}{k} \sum_{j=1}^k \|\tilde{\Phi}(k+1, j)\| = \frac{1}{k} \sum_{j=1}^k \prod_{j=1}^k \left( 1 - \frac{b}{a+j} \right), \quad \mathcal{Z}^* := \limsup_{k \rightarrow \infty} \mathcal{Z}(k) \\ \chi(k) &:= \frac{1}{k} \sum_{j=1}^{k-1} (j^2 \alpha_j^2), \quad \chi^* := \lim_{k \rightarrow \infty} \chi(k). \end{aligned}$$

Then, for all  $i$ , the following holds:

$$\limsup_{k \rightarrow \infty} k \sigma_i^2(k) \leq \frac{1}{N^2} \mathbf{SSNR} + 3 \|S_\eta\| \mathcal{Z}^* + \|S_v\| \chi^*. \quad (59)$$

Moreover, we have:

$$\mathcal{Z}^* \leq \frac{1}{b+1} = \frac{1}{\lambda_2(\mathcal{L})b_0+1} \quad (60)$$

$$\chi^* = b_0^2. \quad (61)$$

*Proof of Lemma 10:* Consider (46). Using the independence of  $\eta(j)$  and  $\eta(k)$ ,  $k \neq j$ , and the independence of  $\eta(k)$  and  $v(j)$ , for all  $k, j$ , and using the equality  $\Phi(k, j) = \tilde{\Phi}(k, j) + J$ , we have:

$$\begin{aligned} \sigma_i^2(k) &= \frac{1}{k^2} \sum_{j=1}^k \text{Var} \left( e_i^\top \Phi(k, j) \eta(j) \right) + \frac{1}{k^2} \sum_{j=1}^{k-1} (\alpha_j j)^2 \text{Var} \left( e_i^\top \Phi(k, j+1) v(j) \right) \\ &= \frac{1}{k^2} k \left( e_i^\top J S_\eta J e_i \right) + \frac{1}{k^2} e_i^\top (I - J) S_\eta (I - J) e_i + \frac{2}{k^2} \sum_{j=1}^{k-1} e_i^\top J S_\eta \tilde{\Phi}(k, j)^\top e_i \\ &\quad + \frac{1}{k^2} \sum_{j=1}^{k-1} e_i^\top \tilde{\Phi}(k, j) S_\eta \tilde{\Phi}(k, j)^\top e_i + \frac{1}{k^2} \sum_{j=1}^{k-1} (\alpha_j j)^2 \left( e_i^\top \Phi(k, j+1) S_v \Phi(k, j+1)^\top e_i \right). \end{aligned}$$

Straightforward algebra shows:

$$e_i^\top J S_\eta J e_i = \frac{1}{N^2} \mathbf{SSNR}.$$

We next bound from above the quantity  $k\sigma_i^2(k)$ , using (V-B) and the following norm arguments: 1)  $\|AB\| \leq \|A\| \|B\|$ ; 2)  $\|Ab\| \leq \|A\| \|b\|$ , for square matrices  $A$  and  $B$ , and a vector  $b$ ; 3)  $\|e_i\| = 1$ ; 4)  $\|\Phi(k, j+1)\| = 1$ . (The latter claim is because  $\|\Phi(k, j+1)\|$  is doubly stochastic.) The bound on  $k\sigma_i^2(k)$  is as follows:

$$\begin{aligned} k\sigma_i^2(k) &\leq \frac{1}{N^2} \mathbf{SSNR} + \frac{2}{k} \|S_\eta\| \sum_{j=1}^{k-1} \|\tilde{\Phi}(k, j)\| \\ &\quad + \frac{1}{k} \|S_\eta\| \sum_{j=1}^{k-1} \|\tilde{\Phi}(k, j)\| + \frac{1}{k} \|S_v\| \sum_{j=1}^{k-1} (\alpha_j j)^2 + \frac{1}{k} e_i^\top (I - J) S_\eta (I - J) e_i \\ &= \frac{1}{N^2} \mathbf{SSNR} + 3\|S_\eta\| \mathcal{Z}(k) + \|S_v\| \chi(k) + \frac{1}{k} e_i^\top (I - J) S_\eta (I - J) e_i. \end{aligned} \quad (62)$$

Taking the lim sup in (62) yields (59).

We now prove (60). Note that  $\mathcal{Z}(k)$  can be written via the following recursion:

$$\begin{aligned} \mathcal{Z}(k+1) &= \left( 1 - \frac{b}{a+k+1} \right) \left( \frac{k}{k+1} \mathcal{Z}(k) + \frac{1}{k+1} \right) \\ \mathcal{Z}(1) &= 1 - \frac{b}{a+1} > 0. \end{aligned}$$

The proof of (60) proceeds analogously to the proof of (51), except that the vector quantity  $(I - J)\epsilon(k)$

is replaced by the scalar  $\mathcal{Z}(k)$ , and the vector  $m_\eta^{(1)}$  is replaced by the scalar 1. The proof of (61) is trivial. Theorem 3 now follows by combining (34) and (35) with (28) to obtain (36). ■

## VI. EXTENSIONS TO NON-GAUSSIAN CASE

We have characterized the exponential decay rate of the error probability in the case of Gaussian (spatially correlated and time-uncorrelated) sensing noise, and Gaussian (spatially correlated and time-uncorrelated) additive communication noise. Our results, to a certain degree, extend to the case when: 1) the zero mean sensing noise is spatially and temporally independent, but with a generic distribution with finite second moments; and 2) the zero mean additive communication noise is spatially correlated, temporally independent, and with a generic distribution and finite second moment. In this case, Theorem 3 remains valid, and all the steps in proving Theorem 3, equations (33)-(35) still go through in the generalized case also (see Appendix.) We now explain the implications of Theorem 3 in the general non-Gaussian model.

Consider  $\mathbf{DSNR}_i(k)$  in (30). In the Gaussian case,  $\mathbf{DSNR}_i(k)$  determines the exponential decay rate of the error probability, as verified by equations (27) and (28). In general, this is no longer the case as higher order moments play a role; however,  $\mathbf{DSNR}_i(k)$  still gives a good estimate for detection performance; see, e.g., [26], [27], [28]. With the optimal centralized detector, we have that:

$$\lim_{k \rightarrow \infty} \frac{\mathbf{DSNR}(k)}{k} = \frac{1}{4} \mathbf{SSNR}.$$

Theorem 3 implies that, with our distributed detector (14), the following holds for all sensors  $i$ :

$$\liminf_{k \rightarrow \infty} \frac{\mathbf{DSNR}_{i,\text{dis}}(k)}{k} \geq \frac{1}{4} \mathbf{SSNR} \frac{\left(1 - \frac{\sqrt{N}}{b_0 \lambda_2(\mathcal{L})} c_\mu\right)^2}{\left(1 + 3 \frac{N}{b_0 \lambda_2(\mathcal{L})} c_\sigma + \frac{N b_0^2}{\mathbf{G}_e}\right)}. \quad (63)$$

Eqn. (63) says that  $\mathbf{DSNR}_{i,\text{dis}}(k)$  grows as  $\Omega(k)$ , as with the optimal centralized detector. This contrasts with  $\mathcal{MD}$  in [5] which achieves only  $\Omega(k^\tau)$ ,  $\tau < 1$ . Second, like before with Theorem 3, now (63) reveals the tradeoff between the information flow and communication noise.

## VII. CONCLUSION

We designed a consensus+innovations distributed detector that achieves exponential decay rate of the detection error probability *at all sensors* under *noisy* communication links, and even when certain (or most sensors) in isolation cannot perform successful detection. This improves over existing work like [5] that achieves a strictly slower rate. We showed how our distributed detector optimally weighs the neighbors' messages via the optimal sequence  $\{\alpha_k\}$ , balancing the two opposing effects: communication



noise and information flow. We found a threshold on the communication noise power above which a sensor that successfully detects the event in isolation still improves its performance through cooperation over noisy links.

## APPENDIX

### A. Proof of Theorem 5

Define first the following quantities:

$$\begin{aligned}\mathcal{Z}_\beta(k) &:= \frac{1}{k} \sum_{j=1}^k \left( \Pi_{s=j}^k (1 - \lambda_N(\mathcal{L})\beta_s)^2 \right), \quad \mathcal{Z}_\beta^* := \lim_{k \rightarrow \infty} \mathcal{Z}_\beta(k) \\ \chi_\beta(k) &:= \frac{1}{k} \sum_{j=1}^{k-1} (j \beta_j)^2, \quad \chi_\beta^* := \lim_{k \rightarrow \infty} \chi_\beta(k).\end{aligned}$$

We will need the following two Lemmas (11 and 12), of which Theorem 5 is a direct corollary.

*Lemma 11* Let Assumptions 1 through 3 hold. In addition, assume Assumption 5. For the weight sequence

$\beta_k = \frac{b_0}{a+k^\tau}$ , we have the following:

$$\mathcal{Z}_\beta^* \begin{cases} = 0 & \text{if } \tau < 1 \\ = 1 & \text{if } \tau > 1 \\ \geq \frac{1}{1+2b_0\lambda_N(\mathcal{L})} & \text{if } \tau = 1 \end{cases} \quad (64)$$

$$\chi_\beta^* \begin{cases} = +\infty & \text{if } \tau < 1 \\ = 0 & \text{if } \tau > 1 \\ = b_0^2 & \text{if } \tau = 1. \end{cases} \quad (65)$$

*Lemma 12*

$$\lim_{k \rightarrow \infty} |\mu_i(k)| = \frac{1}{2} \mathbf{SSNR}_i \quad (66)$$

$$\liminf_{k \rightarrow \infty} k \sigma_i^2(k) \geq \frac{1}{N} \mathbf{SSNR}_i + \mathbf{SSNR}_i \mathcal{Z}_\beta^* + \frac{\lambda_1(S_v)}{N^2} \chi_\beta^*. \quad (67)$$

*Proof of Lemma 11:* We start by proving (64) for  $\tau < 1$ . To this end, note that  $\mathcal{Z}_\beta(k)$  updates according to the following recursion:

$$\begin{aligned}\mathcal{Z}_\beta(k+1) &= \left( 1 - \frac{b'}{a + (k+1)^\tau} \right)^2 \left( \frac{k}{k+1} \mathcal{Z}_\beta(k) + \frac{1}{k+1} \right) \\ \mathcal{Z}_\beta(1) &= \left( 1 - \frac{b'}{a+1} \right)^2 > 0,\end{aligned} \quad (68)$$

where  $b' := b_0 \lambda_N(\mathcal{L})$ . By (68), for sufficiently large  $k_0$ , and for all  $k \geq k_0$ , we have that:

$$\mathcal{Z}_\beta(k+1) \leq \left(1 - \frac{b_\beta}{(k+1)^\tau}\right) \mathcal{Z}_\beta(k) + \frac{1}{k+1},$$

for appropriately chosen  $b_\beta > 0$ . Now, applying Lemma 4 in [29], we get that  $\lim_{k \rightarrow \infty} \mathcal{Z}_\beta(k) = 0$ .

We proceed by proving (64) for  $\tau > 1$ . By (68), and using the fact that  $\mathcal{Z}_\beta(k) \leq 1$ ,  $\forall k$ , the quantity  $\mathcal{Z}_\beta(k+1)$  can be bounded from below as follows:

$$\begin{aligned} \mathcal{Z}_\beta(k+1) &\geq \frac{k}{k+1} \left(1 - \frac{2b'}{a + (k+1)^\tau}\right) \mathcal{Z}_\beta(k) + \left(1 - \frac{2b'}{a + (k+1)^\tau}\right) \frac{1}{k+1} \\ &= \frac{k}{k+1} \mathcal{Z}_\beta(k) - \frac{2b'k}{(k+1)(a + (k+1)^\tau)} \mathcal{Z}_\beta(k) + \left(1 - \frac{2b'}{a + (k+1)^\tau}\right) \frac{1}{k+1} \\ &\geq \frac{k}{k+1} \mathcal{Z}_\beta(k) - \frac{2b''}{k^\tau} + \frac{1}{k+1}, \end{aligned} \quad (69)$$

for appropriately chosen  $b'' > 0$ , and for all  $k \geq k_1$ , where  $k_1$  is sufficiently large. Now, consider the recursion:

$$\begin{aligned} \mathcal{U}(k+1) &= \frac{k}{k+1} \mathcal{U}(k) + \frac{1}{k+1} - \frac{2b''}{k^\tau}, \quad k = k_1, k_1 + 1, \dots \\ \mathcal{U}(k_1) &= \mathcal{Z}_\beta(k_1). \end{aligned} \quad (70)$$

Clearly,  $\mathcal{Z}_\beta(k) \geq \mathcal{U}(k)$ , for all  $k \geq k_1$ . Subtracting 1 from both sides in (70) and applying Lemma 9 yields  $\mathcal{U}(k) \rightarrow 0$ , and, hence,  $\liminf_{k \rightarrow \infty} \mathcal{Z}_\beta(k) \geq 1$ ; on the other hand,  $\mathcal{Z}_\beta(k) \leq 1$ , for all  $k$ , and, hence, (64) for  $\tau > 1$  holds.

To prove (64) and  $\tau = 1$ , consider (69); as  $\tau = 1$ , we have:

$$\mathcal{Z}_\beta(k+1) \geq \frac{k}{k+1} \left(1 - \frac{2b'}{a + k + 1}\right) \mathcal{Z}_\beta(k) + \frac{1}{k+1} - \frac{2b'}{(k+1)^2}, \quad k = 1, 2, \dots$$

Now, define the recursion

$$\mathcal{V}(k+1) = \frac{k}{k+1} \left(1 - \frac{2b'}{a + k + 1}\right) \mathcal{V}(k) + \frac{1}{k+1} - \frac{2b'}{(k+1)^2}, \quad k = 1, 2, \dots$$

Similarly to the proof of (33) in Theorem 3, it can be shown that  $\mathcal{V}(k) \rightarrow \frac{1}{1+2b'}$ . Noting that  $\mathcal{Z}_\beta(k) \geq \mathcal{V}(k)$ ,  $k = 1, 2, \dots$  yields (64) for  $\tau = 1$ .

The proofs of (65) for  $\tau < 1$ ,  $\tau > 1$ , and  $\tau = 1$  are trivial. ■

*Proof of Lemma 12:* Note that, under the assumptions of Lemma 12,  $S_\eta = \text{SSNR}_i I$ . Thus, in (62), the term

$$\frac{2}{k^2} \sum_{j=1}^{k-1} e_i^\top J S_\eta \tilde{\Phi}(k, j)^\top e_i = \frac{16}{k^2} \frac{1}{8} \text{SSNR}_i \sum_{j=1}^{k-1} e_i^\top J \tilde{\Phi}(k, j)^\top e_i = 0, \quad (71)$$

because  $J\tilde{\Phi}(k, j) = 0$ . Multiplying (62) by  $k$ , and using (71), we get:

$$k\sigma_i^2(k) = \frac{1}{N} \text{SSNR}_i + \frac{\text{SSNR}_i}{k} \sum_{j=1}^k e_i^\top \tilde{\Phi}(k, j) \tilde{\Phi}(k, j)^\top e_i + \frac{1}{k} \sum_{j=1}^k (j\beta_j)^2 \left( e_i^\top \Phi(k, j) S_v \Phi(k, j)^\top e_i \right) \quad (72)$$

We next bound  $k\sigma_i^2(k)$  from above, using the following simple relations:

$$\begin{aligned} b^\top A b &\geq \lambda_1(A) \|b\|^2, \quad \|e_i\| = 1 \\ \|\Phi(k, j)^\top e_i\|^2 &\geq \frac{1}{N^2}, \end{aligned} \quad (73)$$

where (73) holds true because  $\Phi(k, j)$  is doubly stochastic. The upper bound on  $k\sigma_i^2(k)$  is as follows:

$$\begin{aligned} k\sigma_i^2(k) &\geq \frac{1}{N} \text{SSNR}_i + \frac{\text{SSNR}_i}{k} \sum_{j=1}^k \lambda_1 \left( \tilde{\Phi}(k, j) \tilde{\Phi}(k, j)^\top \right) + \frac{1}{k} \sum_{j=1}^{k-1} (j\beta_j)^2 \lambda_1(S_v) \|\Phi(k, j+1)^\top e_i\|^2 \\ &\geq \frac{1}{N} \text{SSNR}_i + \frac{\text{SSNR}_i}{k} \sum_{j=1}^k \left( \Pi_{s=j}^k (1 - \lambda_N(\mathcal{L})\beta_s)^2 \right) + \frac{1}{k} \frac{1}{N^2} \sum_{j=1}^{k-1} (j\beta_j)^2 \lambda_1(S_v) \\ &= \frac{1}{N} \text{SSNR}_i + \frac{\text{SSNR}_i}{k} \sum_{j=1}^k \left( \Pi_{s=k-j}^k (1 - \lambda_N(\mathcal{L})\beta_s)^2 \right) + \frac{1}{k} \frac{1}{N^2} \sum_{j=1}^{k-1} (j\beta_j)^2 \lambda_1(S_v) \\ &= \frac{\text{SSNR}_i}{N} + \text{SSNR}_i \mathcal{Z}_\beta(k) + \frac{\lambda_1(S_v)}{N^2} \chi_\beta(k). \end{aligned} \quad (74)$$

Taking the  $\liminf$  in (74) yields (67). ■

### B. Decay rate of the error probability for the $\mathcal{MD}$ algorithm in [5]

We show that, under Gaussian assumptions, with the Algorithm  $\mathcal{MD}$  in ([5], eqn. (14)), the error probability decays to zero at a rate slower than exponential. Recall that  $x_i(k)$ ,  $\mu_i(k)$ ,  $\sigma_i^2(k)$ , and  $P_i^e(k)$  are the sensor  $i$ 's decision variable, its mean and variance under  $H_1$ , and its local error probability, respectively. (Now latter quantities correspond to  $\mathcal{MD}$  and no longer to (12).) Denote by  $P_w^e(k)$  the worst error probability at time  $k$  among sensors:

$$P_w^e(k) = \max_{i=1, \dots, N} P_i^e(k). \quad (75)$$

We show that:<sup>7</sup>

$$P_w^e(k) = \Omega \left( e^{-ck^\tau} \right), \quad 0.5 < \tau < 1, \quad c > 0. \quad (76)$$

Denote by  $x(k)$  the vector of  $x_i(k)$ 's, as before, and  $\Sigma(k) := \text{Cov}(x(k))$ . We will prove (76) by showing:

$$\text{tr}(\Sigma(k)) = \Omega(1/k^\tau). \quad (77)$$

Namely, with  $\mathcal{MD}$ ,  $\lim_{k \rightarrow \infty} \mu_i(k) = \frac{1}{2N} \text{SSNR}$ ,  $\forall i$ . Thus, for all  $k \geq k'$ , for appropriate  $k' > 0$ :

$$\begin{aligned} P_w^e(k) &= \mathcal{Q} \left( \min_{i=1, \dots, N} \frac{\mu_i(k)}{\sigma_i(k)} \right) \geq \mathcal{Q} \left( \frac{\frac{1}{N} \text{SSNR}}{\max_{i=1, \dots, N} \sigma_i(k)} \right) \geq \mathcal{Q} \left( \frac{\frac{1}{N} \text{SSNR}}{\sqrt{\frac{1}{N} \text{tr}(\Sigma(k))}} \right) \\ &\geq \mathcal{Q} \left( \frac{c_p}{\frac{1}{\sqrt{k^\tau}}} \right), \end{aligned} \quad (78)$$

for all  $k \geq k'$  and for appropriately chosen  $c_p > 0$ . Now, applying the upper bound on the  $\mathcal{Q}$  function in (26) to (78) yields (76). It remains to show (77). Denote by  $\mathcal{W}(k) := I - \beta_k L - \alpha_k I$  the updating matrix in the  $\mathcal{MD}$  algorithm, where  $\beta_k = \frac{b}{(k+1)^\tau}$ ,  $\tau \in (\frac{1}{2}, 1)$ , and  $\alpha_k = \frac{a}{(k+1)^\tau}$ ,  $a, b > 0$ . In our notation, the update rule for  $x(k)$  with  $\mathcal{MD}$  is as follows:

$$x(k+1) = \mathcal{W}(k)x(k) + \beta_k v(k) + \alpha_k \eta(k).$$

It can be shown that the covariance matrix  $\Sigma(k) := \text{Cov}(x(k))$  satisfies the following recurrent equation:

$$\Sigma(k+1) = \mathcal{W}(k)\Sigma(k)\mathcal{W}(k)^\top + \alpha_k^2 S_\eta + \beta_k^2 S_v. \quad (79)$$

(Here  $S_\eta$  and  $S_v$  denote respectively the covariance matrix of the innovations  $\eta(k)$  and of the communication noise  $v(k)$ , as before.) Taking the trace in (79) and after algebraic manipulations, we get:

$$\begin{aligned} \text{tr}(\Sigma(k+1)) &\geq \lambda_1(\mathcal{W}(k)^\top \mathcal{W}(k)) \text{tr}(\Sigma(k)) + \alpha_k^2 \text{tr}(S_\eta) + \beta_k^2 \text{tr}(S_v) \\ &\geq (1 - \alpha_k - \beta_k \lambda_N(\mathcal{L}))^2 \text{tr}(\Sigma(k)) + \beta_k^2 \text{tr}(S_v) \\ &\geq (1 - 2(\alpha_k + \beta_k \lambda_N(\mathcal{L}))) \text{tr}(\Sigma(k)) + \beta_k^2 \text{tr}(S_v) \\ &\geq \left( 1 - \frac{c_\Sigma}{(k+1)^\tau} \right) \text{tr}(\Sigma(k)) + \frac{c_v}{(k+1)^{2\tau}}, \end{aligned}$$

<sup>7</sup>By Theorem 3, with (12),  $P_w^e(k) = O(e^{-ck})$ , hence better than  $\mathcal{MD}$ .

for all  $k \geq k_2$  and  $k_2$  sufficiently large, for appropriately chosen  $c_\Sigma, c_v > 0$ . Now, introduce  $\gamma(k) := \text{tr}(\Sigma(k)) (k+1)^\tau$ . Then, we have:

$$\gamma(k+1) \geq \left(1 - \frac{c_\Sigma}{(k+1)^\tau}\right) \gamma(k) + \frac{c_v}{(k+1)^\tau}.$$

Now, consider the sequence  $\mathcal{S}(k)$  that evolves according to the recursion:

$$\mathcal{S}(k+1) = \left(1 - \frac{c_\Sigma}{(k+1)^\tau}\right) \mathcal{S}(k) + \frac{c_v}{(k+1)^\tau}, \quad k = k_2, k_2+1, \dots, \quad \mathcal{S}(k_0) = \gamma(k_0). \quad (80)$$

Clearly,  $\gamma(k) \geq \mathcal{S}(k) \geq 0$ , for all  $k = k_2, k_2+1, \dots$ . It is easy to show that

$$\lim_{k \rightarrow \infty} \mathcal{S}(k) = \frac{c_v}{c_\Sigma}. \quad (81)$$

Namely, subtracting  $\frac{c_v}{c_\Sigma}$  from both sides of equality (80) yields:

$$\left(\mathcal{S}(k+1) - \frac{c_v}{c_\Sigma}\right) = \left(1 - \frac{c_\Sigma}{(k+1)^\tau}\right) \left(\mathcal{S}(k) - \frac{c_v}{c_\Sigma}\right),$$

which in turn implies (81); now, (81) implies that  $\gamma(k) = \Omega(1)$ . Hence, (77) holds.

## REFERENCES

- [1] F. S. Cattivelli and A. H. Sayed, "Distributed detection over adaptive networks using diffusion adaptation," *IEEE Transactions on Signal Processing*, vol. 59, 2011, Digital Object Identifier: 10.1109/TSP.2011.2107902.
- [2] F. S. Cattivelli and A. H. Sayed, "Distributed detection over adaptive networks based on diffusion estimation schemes," in *Proc. IEEE SPAWC '09, 10th IEEE International Workshop on Signal Processing Advances in Wireless Communications*, Perugia, Italy, June 2009, pp. 61–65.
- [3] —, "Diffusion LMS-based detection over adaptive networks," in *Proc. 39th Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, October 2009.
- [4] P. Braca, S. Marano, V. Matta, and P. Willet, "Asymptotic optimality of running consensus in testing binary hypothesis," *IEEE Transactions on Signal Processing*, vol. 58, no. 2, pp. 814–825, February 2010.
- [5] S. Kar, R. Tandon, H. V. Poor, and S. Cui, "Distributed detection in noisy sensor networks," in *Proc. IEEE ISIT 2011, International Symposium on Information Theory*, Saint Petersburg, Russia, August 2011.
- [6] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," August 2008, submitted for publication, 51 pages. [Online]. Available: arXiv:0809.0009v1 [cs.MA]
- [7] F. R. K. Chung, *Spectral Graph Theory*, ser. CBMS Conference Board of the Mathematical Sciences. Providence, RI: American Mathematical Society, 1997, no. 92.
- [8] S. Stankovic, N. Ilic, M. S. Stankovic, and K. H. Johansson, "Distributed change detection based on a consensus algorithm," in *IFAC Workshop on Estimation and Control of Networked Systems (NECSYS)*, Annecy, France, September 2010.
- [9] R. Viswanatan and P. R. Varshney, "Decentralized detection with multiple sensors: Part I—fundamentals," *Proc. IEEE*, vol. 85, pp. 54–63, January 1997.

- [10] J. F. Chamberland and V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 407–416, February 2003.
- [11] J. N. Tsitsiklis, "Decentralized detection," *Adv. Statis. Signal Processing*, vol. 2, pp. 297–344, 1993.
- [12] R. S. Blum, S. A. Kassam, and H. V. Poor, "Decentralized detection with multiple sensors: Part II—advanced topics," *Proc. IEEE*, vol. 85, pp. 64–79, January 1997.
- [13] S. A. Aldosari and J. M. F. Moura, "Detection in sensor networks: the saddlepoint approximation," *IEEE Transactions on Signal Processing*, vol. 55, no. 1, pp. 327–340, January 2007.
- [14] S. Kar, S. A. Aldosari, and J. M. F. Moura, "Topology for distributed inference on graphs," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2609–2613, June 2008.
- [15] S. Kar and J. M. F. Moura, "Consensus based detection in sensor networks: topology design under practical constraints," in *Proc. IEEE Workshop Inf. Theory in Sensor Networks*, Santa Fe, NM, June 2007.
- [16] M. Alanyali, S. Venkatesh, O. Savas, and S. Aeron, "Distributed detection in sensor networks with packet losses and finite capacity links," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4118–4132, July 2004.
- [17] D. Jakovetic, J. Xavier, and J. M. F. Moura, "Weight optimization for consensus algorithms with correlated switching topology," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3788–3801, July 2010.
- [18] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 355–369, January 2009.
- [19] D. Bajovic, D. Jakovetic, J. Xavier, B. Sinopoli, and J. M. F. Moura, "Asymptotic performance of distributed detection over random networks," in *ICASSP 2011, IEEE International Conference on Acoustics, Speech and Signal Processing*, Prague, Czech Republic, May 2011.
- [20] —, "Distributed detection via Gaussian running consensus: large deviations asymptotic analysis," *IEEE Transactions on Signal Processing*, to appear 2011.
- [21] —, "Distributed detection over time varying networks: large deviations analysis," in *44th Annual Allerton Conf. on Comm., Control, and Comp.*, Monticello, IL, September 2010.
- [22] M. Nevelson and R. Hasminskii, *Stochastic Approximation and Recursive Estimation*. Providence, RI: American Mathematical Society, 1973.
- [23] H. L. V. Trees, *Detection, Estimation, and Modulation Theory—Part I—Detection, Estimation, and Linear Modulation Theory*. John Wiley and Sons, 1968.
- [24] S. Kar and J. M. F. Moura, "Ramanujan topologies for decision making in sensor networks," in *44th Annual Allerton Conf. on Comm., Control, and Comp.*, Allerton, IL, Monticello, Sept. 2006, pp. 145–150, invited paper in Sp. Session on Sensor Networks.
- [25] B. T. Polyak, *Introduction to optimization*. Optimization Software Inc., 1987.
- [26] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation and Time Series Analysis*. Addison-Wesley, 1991.
- [27] N. O'Donoghue and J. M. F. Moura, "On the product of independent complex Gaussians," *submitted to IEEE Transactions on Signal Processing*, June 2011.
- [28] C. Lemoine, E. Amador, and P. Besnier, "On the  $k$ -factor estimation for Rician channel simulation in reverberation chamber," *IEEE Trans. on Antennas and Propagation*, no. 3, pp. 1003–1012, March 2011.
- [29] S. Kar and J. M. F. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs," *IEEE Journal on Selected Topics on Signal Processing*, vol. 5, to appear 2011.